

AD-A035 777

ABERDEEN UNIV (SCOTLAND)
ENUMERATION OF LARGE COMBINATORIAL STRUCTURES.(U)
NOV 76 E M WRIGHT

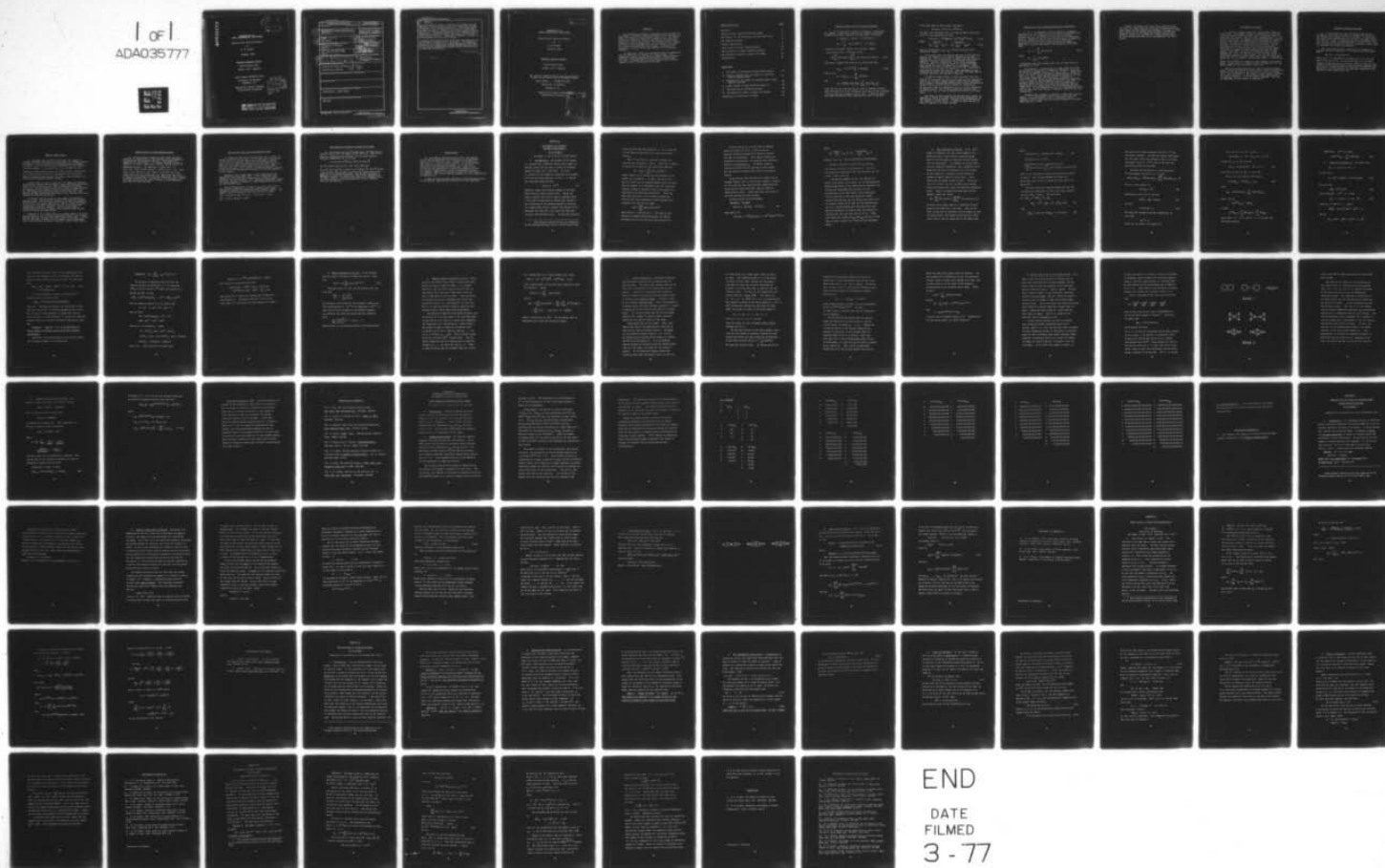
F/G 12/1

UNCLASSIFIED

DA-ERO-75-G-014
NL

1 of 1
ADA035777

1418
1419



END
DATE
FILMED
3-77

ADA035777

(12)

2

AD

ENUMERATION OF
LARGE COMBINATORIAL STRUCTURES

Second Annual Technical Report

by

E. M. Wright

November 1976

EUROPEAN RESEARCH OFFICE

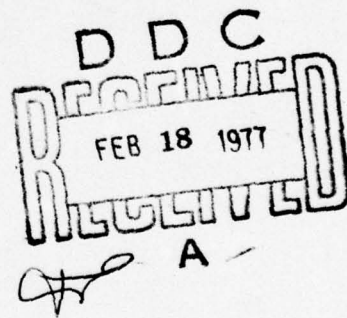
United States Army

London, N.W.1, England

Grant Number DA-ERO-75-G-014

University of Aberdeen
Aberdeen, U.K.

Approved for public release;
distribution unlimited.



COPY AVAILABLE TO DDC DOES NOT
PERMIT FULLY LEGIBLE PRODUCTION

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) ENUMERATION OF LARGE COMBINATORIAL STRUCTURES.		5. TYPE OF REPORT & PERIOD COVERED Final Technical Report November 75 - November 76
7. AUTHOR(s) E.M. WRIGHT		6. PERFORMING ORG. REPORT NUMBER
9. PERFORMING ORGANIZATION NAME AND ADDRESS University of Aberdeen Aberdeen U.K.		8. CONTRACT OR GRANT NUMBER(s) DAERO-75-G-014
11. CONTROLLING OFFICE NAME AND ADDRESS U.S. Army R&S Group (EUR) Box 65 FPO New York 09510		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 61102A 1T161102B18C 00/485
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) Annual technical rept. no. 2 (Final) Nov 75 - Nov 76		12. REPORT DATE November 1976
16. DISTRIBUTION STATEMENT (of this Report) APPROVED FOR PUBLIC RELEASE DISTRIBUTION UNLIMITED		13. NUMBER OF PAGES 87
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		15. SECURITY CLASS. (of this report) UNCLASSIFIED
18. SUPPLEMENTARY NOTES		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) COMBINATORICS; GRAPH THEORY.		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) SEE OVER		

DD FORM 1473

JAN 73

EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

001550 LB

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

20. Abstract

Two methods are described to determine exact formulae for the number of labelled connected graphs on n points and $n+k$ edges for general n and different values of k . One method supplies an essential piece of information for general k ; given this, the other is much more efficient and is well-adapted to machine computation. ~~The author extended the first method to several other problems but has to replace the second by a different method for each problem.~~ According to Temperley, at least two of the problems are of physical interest.

In addition three problems were solved about the properties and the evolution of "almost all" large graphs. The appendices consist of six research papers, three accepted and three under consideration by American and British journals; of these one is by other authors and contains computer results connected with the principal investigator's.

→ these problems ←

* is extended

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

AD

ENUMERATION OF
LARGE COMBINATORIAL STRUCTURES

Second Annual Technical Report

by

E. M. Wright

November 1976

EUROPEAN RESEARCH OFFICE

United States Army

London, N.W.1, England

The research reported herein has been sponsored
in part by the United States Government under

Grant Number : DA-ERO-75-G-014.

University of Aberdeen

Aberdeen, U.K.

Approved for public release
distribution unlimited.

ACCESSION	
NTIS	BASE DATA
DDC	OUT STANDING
UNANNOUNCED	
JUSTIFICATION	
BY	
DISTRIBUTION/STATUS/DATE	
Dist.	ACAD. RES. SPECIAL
A	

Abstract

We describe two methods to determine exact formulae for the number of labelled connected graphs on n points and $n+k$ edges for general n and different values of k . One method supplies an essential piece of information for general k ; given this, the other is much more efficient and is well-adapted to machine computation. We extend the first method to several other problems but have to replace the second by a different method for each problem. According to Temperley, at least two of the problems are of physical interest.

In addition I solve three problems about the properties and the evolution of "almost all" large graphs. The appendices consist of six research papers, three accepted and three under consideration by American and British journals; of these one is by other authors and contains computer results connected with my work.

Table of Contents

Page

Abstract	1
Sparsely-edged connected labelled graphs	3
Equivalence of two generating function identities	5
The reduction method	7
Further applications	9
Large cycles in large labelled graphs	10
The evolution of random unlabelled graphs	11
The degrees of points in almost all graphs	12
Cybernetics?	13

Appendices

1. The number of connected sparsely-edged graphs	14
2. Wright's formulae for the number of connected, sparsely-edged graphs	41
3. Formulae for the number of sparsely-edged strong labelled digraphs	50
4. Large cycles in large labelled graphs II	62
5. The evolution of unlabelled graphs	68
6. The degrees of nodes in almost all graphs	81
References to literature in report	87

Sparsely edged connected labelled graphs

1. As in last year's report, we write $f = f(n, q)$ for the number of connected graphs with n labelled nodes and q edges and without loops or multiple edges. We write

$$N = \frac{1}{2}n(n-1), \quad B(h, k) = h! / \{k!(h-k)!\},$$

$$W_k = \sum_{n=1}^{\infty} f(n, n+k) X^n / n!, \quad D = Xd/dX.$$

I found a recurrence formula for $f(n, n+k)$, namely

$$(n+k+1)f(n, n+k+1) = (N-n+1)f(n, n+k)$$

$$+ \frac{1}{2} \sum_{s=1}^{n-1} B(n, s) s(n-s) \sum_{h=-1}^{k+1} f(s, s+h) f(n-s, n-s+k-1) \quad (1.1)$$

From this I deduced the form of W_{-1} and W_0 and that

$$W_{k+1} = (1-\theta)^{-k-1} \int_{\theta}^1 (1-\theta)^k J_k d\theta, \quad (1.2)$$

where $k \geq 0$,

$$\theta = 1 - DW_{-1} = 1 - \sum_{n=1}^{\infty} n^{n-1} X^n / n!$$

and

$$J_k = \frac{1}{2} \left\{ D^2 W_k - 3DW_k - 2kW_k + \sum_{h=0}^k (DW_h)(DW_{k-h}) \right\},$$

From this we can calculate W_k as a sum of integral (mainly negative) powers of θ for successive k so long as the integrand in (1.2) does not contain a term in θ^{-1} . I found by pencil and paper calculation that this was true for $k = 0, 1, 2, 3$.

I was then able to deduce that, for all n ,

$$2f(n,n) = \{h(n)/n\} - n^{n-2}(n-1)$$

and that, for successive $k \geq 1$ so long as there is no term in θ^{-1} in the integrand, we have

$$f(n,n+k) = (-1)^k M_k h(n) + (-1)^{k-1} P_k n^{n-2}(n-1), \quad (1.3)$$

where

$$M_k = \sum_{s=0}^{[3k/2]-1} m_{ks} n^s, \quad P_k = \sum_{s=0}^{[(3k+1)/2]} p_{ks} n^s \quad (1.4)$$

and the coefficients m_{ks} , p_{ks} can be calculated with sufficient labour.

It is clear that, so long as the term in θ^{-1} does not appear, the formula (1.2) is well adapted for machine calculation. In my last report I wrote that the necessary routines were not available on the Aberdeen University computer. However, since then, three of my colleagues (F. M. D. Gray, A. M. Murray and N. A. Young) have been kind enough to write the appropriate programmes. These evaluate W_k and the m_{ks} and p_{ks} up to $k = 24$. The machine calculations showed that the θ^{-1} term vanished from the integrand up to $k = 24$, at which point the enormous numbers involved saturated the machine's capacity.

In spite of some effort, I could not prove that the term in θ^{-1} vanished for all $k \geq 1$ by means of the recurrence formula (1.2) and it seems unlikely that this is possible. But I found an alternative method, which I call the reduction method, which could be used in theory to calculate W_k .

In practice, however, except for $k = 0, 1$ and 2 , the method is impracticable. But it does enable me to prove that, for all $k \geq 1$, W_k is a sum of powers of θ and so (1.3) must hold.

The whole of these results are written up in detail in Appendix 1, which is a paper which has been accepted by the new Journal of Graph Theory. Appendix 2 contains some of the computer results and is a paper by Gray, Murray and Young to appear in the same journal.

Equivalence of two generating function identities

2. Dr. C. C. Rousseau of the University of Memphis, Tennessee, has been visiting the University of Aberdeen this academic year as a Carnegie Fellow at my invitation. He heard me give two lectures on the work in the preceding section. Somewhat later he came across a paper by Temperley [10] and pointed out to me that a recurrence formula equivalent to (1.2) could be deduced by an extension of Temperley's method from Riddell and Uhlenbeck's [9] result that

$$e^F = 1 + \sum_{n=1}^{\infty} X^n (1+Y)^N / n! \quad (2.1)$$

where

$$F = \sum_{n,q} f(n,q) X^n Y^q.$$

I discovered that the converse was true, so that (2.1) is equivalent to (1.2).

If we write $\mu = (q/n) - \frac{1}{2} \log n$, I have shown [11] that we can deduce from (2.1) an asymptotic expansion of $f(n,q)$ for large q , provided that $\mu \rightarrow \infty$ as $n \rightarrow \infty$. Again from (1.2), we can deduce an asymptotic approximation to $f(n,q)$ when $q-n = O(1)$. Thus the two equivalent results (1.2) and (2.1) about generating functions give us asymptotic information about $f(n,q)$ in two widely separated intervals of q . In part of the intermediate interval, viz. when μ is bounded, Erdős and Rényi [2] have found an asymptotic approximation to $f(n,q)$ in an entirely different way and remark that (2.1) seems useless for this purpose for this range of q . In the remaining part of the intermediate interval, we know only that $f(n,q) = o(B(N,q))$, which is not very informative.

Rousseau and I propose to publish a short note on this topic. It is interesting to note that we now have four methods (including Rényi's [8]) to calculate $f(n,n)$, compared with at least ten [7] to prove Cayley's [1] result that $f(n,n-1) = n^{n-2}$.

Gilbert [5] showed that there were results similar to (2.1) for three types of connected multigraphs and for four types of weakly connected digraphs. For each of these seven cases one can deduce results similar to mine for $f(n, n+k)$ either by the Temperley-Rousseau method from Gilbert's results or directly by a simple adaptation of my first method described in §1 and Appendix 1 of this report.

The reduction method

3. We call a simple graph on n points and q edges an (n,q) graph. We reduce a connected, labelled $(n,n+k)$ graph first by removing every end point and its adjacent edge and repeating this process until we have a smooth graph, i.e. one without end points. We elide a point B of degree 2, adjacent to the edges AB, BC by removing B and replacing AB, BC by an edge AC . We reduce our graph further by eliding all but a carefully defined few of the points of degree 2. Finally we remove the labels. We are left with an unlabelled connected basic $(s,s+k)$ graph, since each step has reduced the number of points and the number of edges by the same amount. The number of such non-isomorphic basic graphs depends only on k . We now write g for the order of the automorphic group of a basic graph, $j(n)$ for the number of connected labelled $(n,n+k)$ graphs which reduce to that basic graph and $J = \sum_n j(n)X^n/n!$. We can show that $J = (1-\Theta)^\alpha \Theta^{-\beta}/g$, where the non-negative integers α and β depend on the basic graph. Clearly $W_k = \sum J$, where the sum is taken over all the basic graphs for the particular value of k .

While this proves that W_k is a sum of integral powers of Θ , it does not provide an efficient, or even a practicable, method of calculating W_k for any but the first few values of k . As k increases, the number of basic graphs increases with great rapidity. It soon becomes quite impracticable to catalogue them all, even by using a computer. All this is explained in more detail in Appendix 1.

Strongly connected digraphs

4. I next found that the reduction method could be applied (in theory) to obtain a formula for the number of various other classes of $(n, n+k)$ graphs. In each case, the method was impracticable for all but a few small values of k but, as in § 3, it supplied essential information about the formula. Using this information, I could in every case but one develop an efficient method (different in each case) to calculate the formula by computer.

The first case was that of $s(n, n+k)$, the number of strongly connected digraphs. The reduction method showed that

$$s(n, n+k) = n! P_k(n) (n > k),$$

where $P_k(n)$ is a polynomial of degree $3k-1$ in n (not the same as the P_k in § 1). To determine $P_k(n)$, it was then enough to find the value of $s(n, n+k)$ for $k < n \leq 4k$ and this could be calculated by means of my recurrence formulae [12] for $s(n, q)$, a simplification of those due to Liskovec [6]. Appendix 3 gives the details and has been submitted to the Bulletin of the London Math. Soc.

Further applications

5. The next case is that of $v(n, n+k)$, the number of "smooth" labelled connected $(n, n+k)$ graphs, i.e. graphs without end-points. Here the reduction method applies very readily; for small k one can take the basic graphs as for the ordinary connected graphs. If we write $V_k = \sum_n v(n, n+k) Z^n / n!$, the reduction method shows us that, for $k \geq 1$, we can find V_k from W_k by substituting $1-Z$ for 0 . The formula for $v(n, n+k)$ in terms of binomial coefficients involving n can be read off at once from the computer results for W_k .

Another case is that of $u(n, n+k)$, the number of labelled blocks with n points and $n+k$ edges. Again the reduction method tells us that $u(n, n+k)$ is $n!$ times a polynomial of degree $3k-1$ in n . But to calculate $u(n, n+k)$ I use a further extension of Temperley's [10] method to find a recurrence formula for U_k , the exponential generating function for $u(n, n+k)$. The formula is more complicated than that for W_k , but is still well adapted for machine calculation.

A further problem is that of the number of Hamiltonian $(n, n+k)$ graphs. The reduction method works again in a rather different way, with the usual limitation on its practicability for all but a few small values of k . So far I have not found a recurrence or other "efficient" method to calculate the exponential generating function. This is still under investigation and none of these three cases has been written up yet in detail for publication.

Temperley [10] remarks that (n, q) graphs, where n is large and q is near n , i.e. $(n, n+k)$ graphs where $k=O(1)$, are of much greater physical interest than those where $q \sim \frac{1}{2}n \log n$. I have not yet had time to follow up this remark, but he does imply that $(n, n+k)$ blocks are also of physical interest when $k=O(1)$.

Large cycles in large labelled graphs

6. I have written a sequel to [14], which has been accepted for publication by the Mathematical Proceedings of the Cambridge Philosophical Society. This appears as Appendix 4 in this report. If a graph contains a cycle of length k for every k such that $3 \leq k \leq k_0$, we say that it is

pancyclic up to k_0 . In Appendix 4 I show that, if $q/n \rightarrow \infty$, $k_0^3 = o(n^2)$ and $k_0^3 = o(q^2/n)$ as $n \rightarrow \infty$, then almost every labelled (n, q) graph is pancyclic up to k_0 .

These conditions are sufficient but, apart from the first, I think it is unlikely that they are necessary. But I cannot at present improve them. The result is of interest in itself but also can be used to obtain a result in the unlabelled case (see the next section).

The evolution of random unlabelled graphs

7. Erdős and Rényi [3,4] showed that, if $q = o(n)$, the typical labelled (n,q) graph (i.e. almost every (n,q) graph) is the union of disjoint trees and so contains no cycle. They trace the elaborate changes in the structure of the typical (n,q) graph as q increases and show that, when q is about $\frac{1}{2}n \log n$ the typical graph consists of one large connected component and a number of isolated points. To this I have added the information contained in [14] and Appendix 4 of this report.

Appendix 5 is a paper which I have submitted to the Journal of the London Math. Soc. In it I show that the evolution of the typical unlabelled (n,q) graph is quite different and much simpler, i.e. simpler to describe, though the result depends on arguments more complicated than those in the labelled case.

The essential point is that, if $q \rightarrow \infty$ as $n \rightarrow \infty$, the typical unlabelled (n,q) graph consists of a large connected component and otherwise only isolated points. The typical graph is pancyclic up to k_0 (i) if $\mu < A$ and $k_0^3 = o(q \log q)$ and (ii) if $\mu \rightarrow \infty$ as $n \rightarrow \infty$, $k_0^3 = (q^2/n)$ and $k_0^3 = o(n^2)$.

The degrees of points in almost all graphs

8. We consider all the 2^N graphs on n labelled points and write $d(x_i)$ for the degree of the point x_i . Then we can show by a probabilistic argument that for some positive C_1, C_2, C_3 independent of n , we have

$$\frac{1}{2}n - C_1(n \log n)^{\frac{1}{2}} < d(x_i) < \frac{1}{2}n + C_1(n \log n)^{\frac{1}{2}}$$

for all points x_i ($1 \leq i \leq n$) and $d(x_i) < \frac{1}{2}n - C_2 n^{\frac{1}{2}}$

for more than $C_3 n / \log n$ points x_i in almost all the 2^N graphs.

I use earlier work of mine to prove that these results are also true for unlabelled graphs. The details are given in Appendix 6, which is a short note that I have submitted to the Math. Annalen. An interesting problem is whether either of these results are best possible and, if not, whether they can be improved.

Cybernetics?

9. The Russian publishing house "Mir" has applied to the Oxford University Press for leave to translate my paper [13] and publish it in a collection of papers on Cybernetics. I have of course agreed. It is interesting that they have selected this paper. It is the most fundamental and the most difficult that I have so far written on the enumeration of unlabelled graphs and from its results I have drawn a number of consequences (see, for example, Appendix 5) and have a number of others still to complete. But I am at present puzzled as to its application to Cybernetics.

I hope to find out more about this in due course, as well as the physical applications referred to at the end of § 5.

Appendix 1

The number of connected sparsely-edged graphs.

E. M. Wright[†]

(To appear in the Journal of Graph Theory)

1. Introduction. We consider (n,q) graphs, i.e. graphs with n labelled points and q edges, no multiple edges and no loops, so that the maximum number of edges is $N = n(n-1)/2$. We write $f = f(n,q)$ for the number of connected (n,q) graphs. It is trivial that $f(n,n+k) = 0$ if $k < -1$ and it has long been known [2] that

$$f(n,n-1) = n^{n-2}. \quad (1)$$

Indeed at least ten different proofs of (1) have been found by various authors [6]. Renyi [7] found a formula for $f(n,n)$, which I discuss later. I have some recollection of having seen (before I was interested in the problem myself) an abstract of a paper which found a formula (not stated in the abstract) for $f(n,n+1)$, but I have not been able to trace this abstract again. In view of the great

[†] The research reported herein was supported by the European Research Office, United States Army.

attention that (1) has attracted, it is curious that no one seems to have taken the study of $f(n, n+k)$ further.

Here I find first a recurrence formula for $f(n, n+k)$ for successive k and n . From this I deduce a recurrence formula for the generating function

$$W_k = W_k(X) = \sum_{n=1}^{\infty} f(n, n+k) X^n / n!,$$

which enables me to express W_k for successive k in terms of a variable θ . In fact, for all $k \geq 1$, W_k is a finite sum of integral powers of θ , but this does not appear to be deducible from the recurrence formula, though of course it can be determined for successive k by detailed calculation. Since W_k does take this form, we can find a formula for $f(n, n+k)$ for each successive k which depends only on powers of n and on the number

$$h(n) = \sum_{s=1}^{n-1} B(n, s) s^s (n-s)^{n-s},$$

where $B(n, s) = n! / [s! (n-s)!]$. All this is well adapted to machine calculation, which is indeed necessary for all but the first few values of k .

To show that W_k is a finite sum of integral powers of θ for all $k \geq 1$, I have to use an entirely different method which does not involve the idea of recurrence. This second method also provides an alternative, but grossly less efficient, method of calculating W_k . In theory, machine calculation could also be applied to this method, but the relative inefficiency would be very great indeed.

If W_k did not take the form of a finite sum of integral powers of θ , we could still deduce a formula for $f(n, n+k)$ but this would be more complicated and would involve additional sums, each at least as complicated as $h(n)$. Since this does not occur, we need not develop the idea further.

My main result is the following.

Theorem 1. We have

$$2f(n, n) = (h(n)/n) - n^{n-2}(n-1) \quad (2)$$

and, for $k \geq 1$,

$$f(n, n+k) = (-1)^k M_k(n) h(n) + (-1)^{k-1} P(n) n^{n-2}(n-1),$$

where

$$M_k = \sum_{s=0}^{[3k/2]-1} m_{ks} n^s, \quad P_k = \sum_{s=0}^{[(3k+1)/2]} p_{ks} n^s$$

and the m_{ks}, p_{ks} can be calculated in succession.

Machine calculation shows that m_{ks} and p_{ks} are positive for all $k \leq 24$ and all s . It is a very plausible conjecture that this persists for all k but I cannot prove it.

I am deeply indebted to Dr. A.M. Murray, Dr. P.M.D. Gray and Dr. N.A. Young, all of the Computing Science Department of the University of Aberdeen, for writing the programmes and carrying out the calculations on the PDP 11/40 in that department. At the stage when I had only found (by "hand" calculation) that W_1, W_2, W_3 and W_4 were each a sum of integral powers of θ , they ran the calculations up to $k = 24$ and showed that this was still true. This encouraged me to persist with the problem of proving that this was true for all $k \geq 1$. They also found the values of m_{ks} and p_{ks} for all $k \leq 24$. Some of their results are listed in the following paper.

2. The recurrence relations. If an (n, q) graph is connected, an edge can be added in $N-q$ different ways to make another connected graph. Thus we form $(N-q)f(n, q)$ connected $(n, q+1)$ graphs (not all different). If instead the (n, q) graph has two connected components, one a connected (s, t) graph and the other a connected $(n-s, q-t)$ graph, we can construct a connected $(n, q+1)$ graph by adding an edge in each of $s(n-s)$ ways. The s points can be chosen in $B(n, s)$ ways. If we sum over all s from 1 to $n-1$, we shall have counted each (n, q) graph with a pair of connected components twice. Hence we construct Q connected $(n, q+1)$ graphs in this way, where

$$2Q = \sum_{s=1}^{n-1} B(n, s) s(n-s) \sum_{t=s-1}^{q-n+s+1} f(s, t) f(n-s, q-t).$$

No other (n, q) graph leads to a connected $(n, q+1)$ graph by the addition of one edge. But, in the total collection of connected $(n, q+1)$ graphs we have now constructed, each graph occurs just $q+1$ times, since each of the $q+1$ edges may be the added edge.

Hence

$$(q+1)f(n, q+1) = (N-q)f(n, q) + Q(n, q). \quad (3)$$

Putting $q=n+k$, we have

$$2(n+k+1)f(n, n+k+1) = 2(N-n-k)f(n, n+k) + \sum_{s=1}^{n-1} B(n, s)s(n-s) \sum_{h=-1}^{k+1} f(s, s+h)f(n-s, n-s+k-h). \quad (4)$$

This is the recurrence formula for $f(n, n+k)$ referred to above. Bol [1] and Dziobek [3] found the particular case of (4) in which $k=-2$ and used it to prove (1).

We write D for the operator Xd/dX and $[V]_n$ for $n!$ times the coefficient of X^n in the power series V . Clearly $[DV]_n = n[V]_n$. We write also $G = DW_{-1}$, $H = D^2W_{-1}$, so that

$$[G]_n = n^{n-1}, [H]_n = n^n, [H^2]_n = h(n) \quad (5)$$

and

$$[W_k]_n = f(n, n+k), [DW_k]_n = n f(n, n+k). \quad (6)$$

The series for G and H converge for $|X| < e^{-1}$ by Stirling's formula. We need not assume convergence for the power series W_k (though in fact it is also convergent when $|X| < e^{-1}$), but we can take the generating function as formal.

Dividing (4) through by $n!$, multiplying by X^n and summing from $n=1$ to ∞ , we have

$$2(D+k+1)W_{k+1} = (D^2 - 3D - 2k)W_k + \sum_{h=-1}^{k+1} (DW_h)(DW_{k-h}). \quad (7)$$

If $k=-2$, this reduces to

$$2(G - W_{-1}) = G^2. \quad (8)$$

Operating on this with D, we have

$$D(G^2) = 2HG = 2(H-G), \quad (9)$$

so that

$$(1-G)(1+H) = 1. \quad (10)$$

Dividing (9) through by XG and integrating, we find that

$$Ge^{-G} = X,$$

which is, of course, well-known [5].

If we put $k=-1$ in (7), we have

$$2(1-G)DW_0 = H - 3G + 2W_{-1} = H - G - G^2$$

by (8) and, so by (9) and (10),

$$2DW_0 = (1+H)(H-G-G^2) = H^2 - H + G \quad (11)$$

From this and (5) and (6), we have (2).

For $k \geq 0$, the equation (7) becomes

$$(1-G)DW_{k+1} + (k+1)W_{k+1} = J_k, \quad (12)$$

where

$$2J_k = (D^2 - 3D - 2k)W_k + \sum_{h=0}^k (DW_h)(DW_{k-h}).$$

Hence, by (9),

$$D(G^{k+1}W_{k+1}) = HG^k J_k,$$

so that

$$G^{k+1}W_{k+1} = \int_0^X HG^k J_k dX/X = \int_0^G G^k J_k dG,$$

since $G(0) = 0$. If we put $\Theta = 1 - G$, we have the following theorem.

Theorem 2. If $k \geq 0$, then

$$(1-\theta)^{k+1}W_{k+1} = \int_{\theta}^1 (1-\theta)^k J_k d\theta. \quad (13)$$

3. Deduction of Theorem 1. We remark that,

$$DG = H = G/(1-G) = \theta^{-1} - 1$$

by (9) and so

$$D \equiv (\theta^{-1} - 1)(d/dG) \equiv (1 - \theta^{-1})(d/d\theta). \quad (14)$$

Now, by (11),

$$2DW_0 = \theta^{-2}(1-\theta)^3, \quad (15)$$

and so $2dW_0/d\theta = -\theta^{-1} + 2 - \theta$ and

$$4W_0 = -2 \log \theta - 3 + 4\theta - \theta^2, \quad (16)$$

since $W_0 = 0$ when $\theta = 1$. Again

$$2D^2W_0 = (1-\theta^{-1})(-2\theta^{-3} + 3\theta^{-2} - 1)$$

and so

$$8J_0 = 5\theta^{-4} - 16\theta^{-3} + 15\theta^{-2} - 5 + \theta^2.$$

Thus, for $k=0$, the term in θ^{-1} in the integrand on the right of (13) vanishes and so the integral is a sum of powers of θ and contains no $\log \theta$ term. We then find that

$$\begin{aligned} 24W_1 &= 5\theta^{-3} - 19\theta^{-2} + 26\theta^{-1} - 14 + \theta + \theta^2 \\ &= \theta^{-3}(1-\theta)^4(5+\theta). \end{aligned} \quad (17)$$

If we now calculate $(1-\theta)J_1$, the term in θ^{-1} vanishes and we can deduce that

$$48W_2 = \theta^{-6}(1-\theta)^4(15-5\theta-2\theta^2-5\theta^3-\theta^4)$$

from (13). Machine calculations for successive k show that the term in θ^{-1} in $(1-\theta)^k J_k$ disappears up to $k=23$, but it does not seem possible to deduce that this is true for all $k \geq 1$ from Theorem 2. I prove the following theorem later by an entirely different combinatorial argument.

Theorem 3. For all $k \geq 1$, W_k is the sum of a finite number of integral powers of θ with rational coefficients.

From (14), (15) and Theorems 2 and 3 we can deduce the following theorem in an obvious way.

Theorem 4. $W_k = \sum_{s=-3k}^2 c_{ks} \theta^s$ if $k \geq 1$.

We now have to determine $f(n, n+k)$ from the form of W_k given in Theorem 4, i.e. to calculate $[\theta^s]_n$ in terms of $h(n)$, n^{n-2} and other powers of n .

By (5) and (9), we have

$$[G^2]_n = 2n^{n-2}(n-1), [\theta]_n = -n^{n-1}, [\theta^2]_n = -2n^{n-2}.$$

For the negative powers of θ , we remark that

$$H = \theta^{-1} - 1, H^2 = \theta^{-2} - 2\theta^{-1} + 1$$

and, by (14),

$$DH = (1 - \theta^{-1})(dH/d\theta) = \theta^{-3} - \theta^{-2},$$

$$DH^2 = 2\theta^{-4} - 4\theta^{-3} + 2\theta^{-2},$$

and so on, in succession. Hence

$$n^n = [\theta^{-1}]_n, h(n) = [\theta^{-2} - 2\theta^{-1}]_n,$$

$$[\theta^{-2}]_n = h(n) + 2n^n, [\theta^{-3}]_n = h(n) + n^n(n+2),$$

$$2[\theta^{-4}]_n = (n+2)h(n) + 4n^n(n+1)$$

and so on. Thus from (17) we deduce that

$$24f(n, n+1) = n^{n-2}(n-1)(5n^2+3n+2) - 14h(n)$$

and similarly we can calculate that

$$\begin{aligned} 1152f(n, n+2) &= (45n^2 + 386n + 312) h(n) \\ &- 4n^{n-2}(n-1)(55n^3 + 36n^2 + 18n + 12). \end{aligned}$$

More generally we deduce the second part of Theorem 1 from Theorem 4. Obviously $f(1, 1+k) = 0$ if $k \geq 1$; since $h(1) = 0$, the factor $n-1$ must appear as it does.

4. Renyi's formula for $f(n,n)$. As we remarked earlier, Renyi [7] found a formula for $f(n,n)$. This was

$$f(n,n) = \frac{1}{2} \sum_{s=3}^n s! B(n,s) n^{n-s-1}. \quad (18)$$

The equivalence of (18) and (2) follows from the fact that

$$\frac{h(n)}{n!} = \sum_{s=2}^n \frac{n^{n-s}}{(n-s)!}$$

We can prove this either by using Cauchy's theorem and the transformation $X = Ge^{-G}$ or by expanding $(n-s)^{n-s}$ in each term of $h(n)$ by means of the binomial theorem, re-arranging the resulting double sum and remarking that

$$\sum_{s=0}^v \frac{(-1)^{v-s} s^v}{s!(v-s)!} = S(v,v) = 1,$$

where $S(n,k)$ is the Stirling number of the second kind.

5. Another method to calculate $f(n,n)$. Since at least ten authors have found as many proofs (see [6]) of (1) between 1889 and 1963, it is perhaps worthwhile to find still another proof of (2), so that Renyi and I will have found three proofs of (2) or (18) between 1951 and 1975. More seriously, my second proof gives a combinatorial reason why W_0 involves $\log \theta$, i.e. why Theorem 3 is only true for $k \geq 1$, and also helps to make my proof of that theorem in §6 a little more easily understood.

Any connected labelled (n,n) graph contains just one circuit of t points, where $t \geq 3$, and each point in the circuit is a root of a rooted tree (possibly the degenerate tree consisting of a single point.) The number of ways in which we can choose first s_1 points out of the n , then s_2 , then s_3 , ..., up to s_t is $n!/(s_1! s_2! \dots s_t!)$. The s_1 points in the first set can form $s_1^{s_1-1}$ rooted trees. Let \sum' denote summation over all ordered sets of positive integers s_1, \dots, s_t such that $\sum s_i = n$. Then n labelled points can be arranged round the circuit

in t rooted trees in $n! \psi(n, t)/(2t)$ ways, where

$$\psi(n, t) = \sum' s_1^{s_1-1} s_2^{s_2-1} \dots s_t^{s_t-1} / (s_1! \dots s_t!),$$

since each possible circuit has been counted $2t$ times in $n! \psi(n, t)$. Hence

$$f(n, n) = n! \sum_{t=3}^n \psi(n, t)/(2t)$$

and so

$$\begin{aligned} 2W_0 &= 2 \sum_{n=1}^{\infty} f(n, n) X^n / n! = \sum_{t=3}^{\infty} \left(\sum_{s=1}^{\infty} s^{s-1} X^s / s! \right)^t / t \\ &= \sum_{t=3}^{\infty} G^t / t = -\log(1-G) - G - (G^2/2), \end{aligned}$$

which is equivalent to (16). We can deduce (11) by operating with D and (2) follows as before.

6. Proof of Theorem 3. From each connected $(n, n+k)$ graph we remove every end-point and its adjacent edge. We repeat this process until we are left with a connected $(t, t+k)$ graph with no end-points. Since $k \geq 1$, this graph will contain a positive number r of points of degree greater than 2, which we call special points. Clearly $r \leq 2k$. We now consider special paths. Such a path begins at a special point, has only ordinary points (i.e. points of degree 2) in its interior and ends at a special point. It is easily seen that the $(t, t+k)$ graph consists of a number of special paths, disjoint except for the special points at the ends.

These paths are of four kinds. An α -path begins and ends at the same special point and so must have at least 2 interior points. We elide all its interior points except 2, i.e. if B is an ordinary point and AB, BC are edges, we replace AB, BC by AC and delete B . If two different special points are joined by just one special path, this is a β -path; we elide all its interior points. If two different special points are joined by more than one special path, at most one

of these paths is a single edge, which we call a δ -path. The remaining paths, or all the paths if there is no δ -path, joining the two points each have at least one interior point and are χ -paths; we elide from each χ -path all but one interior point. We call the new graph (with its labels removed) a basic graph. The number of α -, β -, χ -, δ -paths is a , b , c , d respectively. The number of points in the basic graph is $r + 2a + c$. Each elision has removed one point and one edge. Hence the number of edges in the basic graph is

$$3a + b + 2c + d = r + 2a + c + k$$

and so $a + b + c + d = r + k \leq 3k$.

Thus the number of non-isomorphic basic graphs depends only on k .

We now take any one of the basic graphs, write $j(n)$ for the number of connected labelled $(n, n+k)$ graphs from which the basic graph can be obtained by the above process and put $J = \sum_n j(n) X^n / n!$. We label the special paths. To obtain any of our

original $(t, t+k)$ graphs without end points we distribute $t-r-2a-c$ additional nodes on the $a+b+c$ paths which are α -, β - or γ - paths. We can do this in just $y=y(t) = y(t, n, a, b, c)$ different ways, where y is the number of non-negative integral solutions of

$$m_1 + \dots + m_{a+b+c} = t-r-2a-c,$$

order being relevant, so that y is the coefficient of G^t in $G^{r+2a+c}(1-G)^{-a-b-c}$. In fact, $y = B(t-r-a+b-1, a+b+c-1)$, but this is irrelevant to our argument.

We now discard the labels from the special paths and label all t of the points in our new $(t, t+k)$ graph, as (say) T_1, \dots, T_t . Taking the n labelled points in the $(n, n+k)$ graph we are seeking to construct, we can arrange them in t ordered rooted trees in $n! \psi(n, t)$ ways, and root each tree at the corresponding point in the $(t, t+k)$ graph, i.e. the tree on the first s_1 points at T_1 , and so on. Thus we have constructed $y(t)n! \psi(n, t)$ of the $(n, n+k)$ graphs from each of

which the same basic graph could be obtained. All such graphs are included but not all are different. In fact, each of these latter graphs will occur just g times, where g is the order of the graph of automorphisms of our original basic graph. Hence we have

$$j(n) = \sum_{t \leq n} n! y(t) \psi(t, n) / g$$

and so

$$gJ = \sum_t y(t) G^t = G^{r+2a+c} (1-G)^{-a-b-c}$$

and

$$J = (1-\theta)^{r+2a+c} \theta^{-a-b-c} / g,$$

a finite sum of integral powers of θ . Summing over all the basic graphs, we obtain Theorem 3.

7. Calculation of W_k by the second method. If we

wish to use the second method to calculate W_k , we must first determine all the basic graphs. We can reduce these further by eliding all the remaining points interior to the α - and γ - paths. We have then a smaller number of non-isomorphic connected unlabelled multigraphs, each on r points of degree 3 or more, with $r+k$ edges, loops and multiple edges being allowed. If we now remove the loops and replace every bunch of multiple edges by a single edge, we have a connected simple graph on r nodes and not more than $r+k$ edges. For $k \leq 3$, these may be read off from the diagrams in [4].

In the reverse process, we have to consider each connected simple unlabelled graph with r points, where $1 \leq r \leq 2k$, and not more than $r+k$ edges. To each fundamental graph, we determine all possible ways (if any) of adding loops and/or multiple edges parallel to existing ones so as to make all points of degree at least 3 and not to exceed a total of $r+k$ edges. If we still have edges to spare, i.e.

we have not used all $r+k$ edges, we have to consider all possible ways of adding the remaining edges as loops or multiple edges parallel to existing ones. We thus obtain all our multigraphs (as in Figure 1 for $k = 1$). From these we can determine our basic graphs quite simply (as in Figure 2 for $k = 1$).

By the reasoning of the last section we find that

$$w_1 = \frac{(1-\theta)^5}{8\theta^2} + \frac{(1-\theta)^6}{8\theta^3} + \frac{(1-\theta)^5}{12\theta^3} + \frac{(1-\theta)^4}{4\theta^2},$$

each of the terms on the right corresponding to one of the basic graphs in Figure 2. From this we deduce that

$$24w_1 = \theta^{-3}(1-\theta)^4(5+\theta),$$

which agrees with (17).

If $k = 2$, we find 15 multigraphs and 27 basic graphs. At this stage, it is natural to conjecture (very tentatively) that we may expect $1.3.5 \dots (2k+1)$ multigraphs and $(k+1)^{k+1}$ basic graphs for every k , but this is false for $k = 3$ (but only by 2 in each case), since we have 107 multigraphs and 254 basic graphs, instead of 105 and 256. For $k = 4$, we may

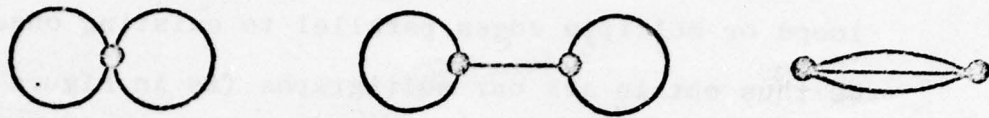


FIGURE 1

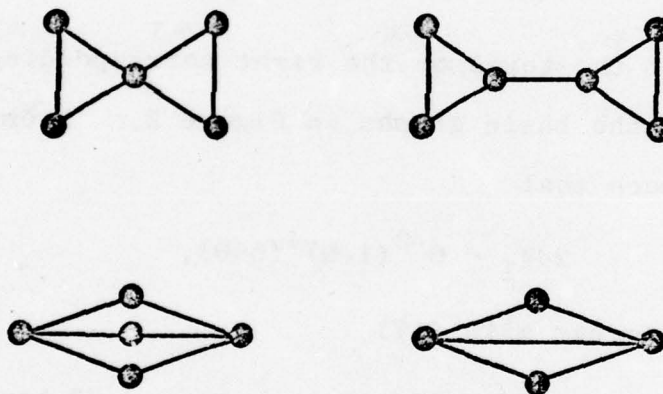


FIGURE 2

expect some 900 to 1000 multigraphs and about 3000 basic graphs.

The bulk of the labour of calculating W_k by this method lies in constructing the multigraphs (and, for $r > 6$, i.e. for $k > 3$, Harary's diagrams in [4] would not give us all the fundamental connected graphs). The subsequent construction of the basic graphs and the calculations to deduce W_k are relatively simple. I have calculated W_2 from the 27 basic graphs. But, for $k > 3$, the method clearly becomes impossibly laborious and increasingly inefficient in comparison with the first method. One could presumably write a programme for a machine to use this method for a few further k , but the relative inefficiency would still be large. In fact, the interest of the second method seems to lie almost entirely in its use to prove Theorem 3. Less importantly, we now have two different methods to determine $f(n, n+1)$ and $f(n, n+2)$, compared to the three for $f(n, n)$ and the ten or more for $f(n, n-1)$.

8. Asymptotic results for large n . In a sequel we shall show that, for large n , we have

$$h(n) \sim n!e^{n/2} \sim \sqrt{(\pi n/2)} n^n$$

and from this and (2) we deduce that

$$f(n,n) \sim \sqrt{(\pi/8n)} n^n$$

in agreement with Renyi [7]. More precisely, we can use a result of [8] to show that

$$h(n) = (n!e^{n/2}) - Tn^n,$$

where

$$T = \frac{4}{3} + \frac{4}{135n} - \frac{8}{2835n^2} - \frac{97}{45360n^3} - \frac{61399}{119750400n^4} + O\left(\frac{1}{n^5}\right),$$

and more terms can be calculated if required. This can be used in the formula of Theorem 1 to find an asymptotic expansion for $f(n,n+k)$.

Using (10) in (12), we have

$$DW_{k+1} + (k+1)(1+H)W_{k+1} = (1+H)J_k. \quad (19)$$

If $k=O(1)$ as $n \rightarrow \infty$, we can use the Euler-Maclaurin sum formula to deduce directly from (19) that

$$f(n, n+k) = \varrho_k n^{n+[(3k-1)/2]} \left\{ 1 + O(n^{-\frac{1}{2}}) \right\},$$

where

$$\varrho_k = \pi^{\frac{1}{2}} 2^{(1-3k)/2} \sigma_k / \Gamma((3k/2) + 1),$$

$$4\sigma_0 = 1, \quad 16\sigma_1 = 5, \quad 16\sigma_2 = 15,$$

$$\sigma_{k+1} = \left\{ 3(k+1)\sigma_k/2 \right\} + \sum_{s=1}^{k-1} \sigma_s \sigma_{k-s} \quad (k \geq 2).$$

Note added August 14, 1976. The second (reduction) method can be extended to find formulae for general n for the number of connected labelled $(n, n+k)$ graphs which are (i) without end points or (ii) blocks or (iii) Hamiltonian and for the number of strongly connected labelled $(n, n+k)$ digraphs. In all four cases the method becomes impracticable for k greater than 2 or 3. For all except the Hamiltonian case, however, the reduction method supplies us with an essential piece of information (different in each case) for general k and this enable us to find another method, again entirely different in each case, which is well adapted for machine calculation and which is practicable for rather larger k .

References in Appendix 1

- [1] G. Bol, Über eine kombinatorische Frage.
Abh. Math. Sem. Hansische Univ. 12(1938), 242-245.
- [2] A. Cayley, A theorem on trees. Quart. J. Math.
23(1889), 376-378.
- [3] O. Dziobek, Eine Formel der Substitutionstheorie.
Sitz, Berliner Math. Ges. 17(1917), 64-67.
- [4] F. Harary, Graph theory. Addison-Wesley, Reading
Mass. (1969), 215-222.
- [5] A. Hurwitz and R. Courant, Funktionentheorie.
Springer, Berlin. 3rd ed. (1929), 141-142.
- [6] J. W. Moon, Various proofs of Cayley's formula for
counting trees, A seminar on graph theory. (ed. F. Harary)
New York 1967, 70-78.
- [7] A. Rényi, On connected graphs I, Publ. Math. Inst.
Hungarian Acad. Sci. 4 (1959), 385-388.
- [8] E. M. Wright, Solution of the equation $ze^z = a$.
Proc. Roy. Soc. Edinburgh. A65(1959), 193-203.

Appendix 2
Wright's formulae for the number
of connected sparsely-edged graphs

P.M.D. Gray, A.M. Murray and N.A. Young

(To appear in the Journal of Graph Theory)

1. Introduction. Computer programs have been written using the recurrence relations obtained by Wright [1] for the calculation of $f(n, n+k)$ from the formulae of his Theorem 1. Exact values have been obtained for the coefficients m_{ks} and p_{ks} for all k up to $k = 24$. All of these coefficients are positive. Values up to $k = 8$ are given in this paper.

2. Computational Method. Two separate computer programs have been written for this problem. One in PDP11 Macro (A.M.M) which is described herein obtained values for the coefficients m_{ks} and p_{ks} up to $k = 24$. The other in PDP11 Fortran (P.M.D.G and N.A.Y) which used slightly different algorithms checked these results up to $k = 12$. Both programs were run on the PDP11/40 in the Department of Computing Science.

The formulae used in the programs to obtain the W_k are derived from Wright's equations (12) and (13). The J_k and W_{k+1} are obtained in succession as series of positive and negative powers of θ , using as initial value the series

for DW_0 in (15). The vanishing of the coefficient of θ^{-1} in the integrand of (13) at each stage provides a check on the arithmetic.

Using Wright's (5) and (9) it can be shown that $(-1)^{k_{m_{ks}}}$ and $(-1)^{k_{p_{ks}}}$ are the coefficients of $D^S H^2$ and $D^S G^2/2$ respectively in W_k , all expressed as power series in θ . The m_{ks} and p_{ks} are obtained by successively eliminating multiples of $D^S H^2$ and $D^S G^2/2$ from W_k , starting with the series containing the lowest power of θ . The method of calculating the m_{ks} and p_{ks} provides an effective check on the arithmetic. Since the number of coefficients of θ in each W_k exceeds by two the number of $D^S H^2$ and $D^S G^2/2$ involved, the remaining two coefficients must vanish.

The numbers involved in the calculations are rational fractions, the numerators of which increase rapidly with k reaching 10^{106} for $k = 24$. Each rational fraction is represented by a sign, a positive integer stored in multiple length binary, and a sequence of signed exponents of primes. Sufficient primes are used for each fraction to include all those which occur in the denominator. For each k , only primes less than $3k$ are required. All factors of the primes used are extracted from both the numerator and

denominator. The positive integer in the representation is the factor of the numerator which remains after extracting products of primes. All numbers which are used as divisors can be expressed entirely as a product of positive or negative powers of the primes used.

Using this representation, division is done by subtracting exponents and multiplication by adding exponents and taking the product of the positive integers. Addition and subtraction are done by taking the smallest power of each prime as a common factor and expanding the remainder of each operand. After adding or subtracting these using multiple length arithmetic, the result is reduced to standard form by simplification and factorisation.

3. Results

s	$12m_{1s}$	s	$24p_{1s}$
0	7	0	2
		1	3
		2	5

s	$48 \times 4! m_{2s}$	s	$12 \times 4! p_{2s}$
0	312	0	12
1	336	1	18
2	45	2	36
		3	55

s	$144 \times 7! m_{3s}$	s	$72 \times 7! p_{3s}$
0	126,000	0	9,576
1	225,588	1	14,490
2	119,140	2	31,815
3	33,075	3	67,143
		4	31,831
		5	3,315

s	$384 \times 10! m_{4s}$	s	$96 \times 10! p_{4s}$
0	176,843,520	0	6,531,840
1	377,944,416	1	9,979,200
2	291,981,480	2	23,375,520
3	140,233,100	3	57,670,128
4	35,976,150	4	43,193,532
5	1,601,775	5	14,025,040
		6	3,178,875

s	$192 \times 13! m_{5s}$	s	$48 \times 13! p_{5s}$
0	119,143,664,640	0	4,265,015,040
1	284,416,073,856	1	6,575,849,280
2	267,988,601,088	2	16,193,136,960
3	166,315,137,560	3	44,130,297,168
4	67,964,819,780	4	41,375,851,056
5	13,958,634,690	5	19,578,697,276
6	1,603,376,775	6	7,111,163,740
		7	1,323,797,475
		8	59,634,000

s	$288/5 \times 16! m_{6s}$	s	$72/5 \times 16! p_{6s}$
0	98,647,633,244,160	0	3,424,861,440,000
1	254,676,192,634,368	1	5,325,140,620,800
2	272,652,031,140,480	2	13,660,215,528,960
3	197,892,051,026,176	3	40,000,445,276,544
4	102,197,036,781,840	4	43,171,991,234,496
5	32,338,292,146,160	5	25,167,812,651,168
6	6,919,958,485,440	6	11,632,130,580,792
7	904,499,095,500	7	3,535,865,786,860
8	23,929,030,125	8	564,608,222,850
		9	62,153,240,625

s 672×19!m_{7s}

Ø 5,668,999,582,108,262,400

1 15,535,467,573,722,772,480

2 18,223,444,759,662,505,728

3 14,743,403,769,011,791,104

4 8,833,912,970,481,602,080

5 3,534,868,345,357,199,760

6 1,025,141,485,305,938,000

7 218,999,364,535,551,600

8 26,564,355,496,428,750

9 1,893,623,798,941,875

s 168×19!p_{7s}

Ø 191,213,708,063,938,560

1 299,607,142,277,963,520

2 795,607,706,229,336,000

3 2,463,144,465,639,795,840

4 2,935,902,159,828,005,904

5 1,964,209,983,197,753,808

6 1,058,708,432,388,650,048

7 412,939,628,429,328,000

8 100,686,761,448,521,920

9 18,714,587,986,713,360

10 2,077,737,705,302,175

11 58,152,949,974,000

s	12288/7×22!m _{8s}	s	3072/7×22!p _{8s}
0	118,575,433,027,180,678,348,800	0	3,893,205,483,768,958,156,800
1	340,821,867,074,982,721,781,760	1	6,143,328,069,445,376,409,600
2	428,704,644,018,874,371,158,016	2	16,810,155,154,980,207,820,800
3	376,150,948,331,993,348,760,576	3	54,448,058,107,681,542,881,280
4	250,470,092,279,344,676,471,040	4	69,964,261,525,350,589,667,328
5	116,903,504,037,607,169,734,720	5	51,722,568,064,934,246,131,200
6	41,003,229,834,970,246,500,800	6	31,098,929,223,562,058,723,712
7	11,312,762,332,522,028,378,000	7	14,227,334,318,390,328,825,440
8	2,148,623,401,492,125,558,000	8	4,405,206,715,317,464,185,840
9	288,608,522,299,040,443,500	9	1,079,307,694,362,844,246,960
10	25,266,853,210,026,928,500	10	196,956,914,112,258,801,600
11	475,694,276,957,274,375	11	20,373,506,810,219,697,750
		12	1,479,022,312,282,081,875

4. Acknowledgement. We are grateful to E.M. Wright for bringing this interesting computational problem to our attention.

Reference in Appendix 2

1. E.M. Wright, "The number of connected sparsely-edged graphs", submitted to the Journal of Graph Theory.

Appendix 3

Formulae for the number of sparsely-edged strong labelled digraphs

E. M. Wright †

(Submitted to the Bulletin of the London Math. Soc.)

1. Introduction. In a digraph each pair of different points A,B may be joined by the directed edge AB, or by the directed edge BA or by both or by neither. An (n,q) digraph has n points and q directed edges. A digraph is strong (or strongly-connected) if there is a directed path from each point to every other point. We write $s(n,q)$ for the number of strong labelled (n,q) digraphs. It is trivial that $s(n,n+k) = 0$ if $k < 0$ and almost trivial that $s(n,n) = (n-1)!$. I shall prove the following theorem.

Theorem. If $n > k > 0$, then

$$s(n,n+k) = n!P_k(n),$$

where $P_k(n)$ is a polynomial in n of degree $3k-1$.

In particular, $P_1(n) = \frac{1}{4}(n-2)(n+3)$.

† The research reported herein was supported by the European Research Office of the United States Army.

The method of proof can be used with some labour to determine the coefficients in $P_2(n)$, but becomes impossibly laborious for $k \geq 3$. However my results in [2], which were a simplification of Liskovec's [1], are well adapted for machine computation and enable one to calculate the values of $s(n, n+k)$ for $n \leq 4k$ and small values of k and so the value of $P_k(n)$ for $k+1 \leq n \leq 4k$. From these the coefficients of $P_k(n)$ can be determined.

We put $0! = 1$ and write

$$B(h, m) = h! / \{ m!(h-m)! \} \text{ if } h \geq m \text{ and } B(h, m) = 0 \text{ if } h < m.$$

2. Proof of first part of Theorem. My method is an adaptation of the one I used [3] to determine the form of $f(n, n+k)$, the number of connected graphs on n points and $n+k$ edges. As in [3], it is an inefficient method to determine the actual formula for $s(n, n+k)$ for a given k , unless k is small. In [3] I also calculated the number of ways in which a connected $(n, q+1)$ graph could be formed from an (n, q) graph by the addition of a new edge and deduced a recurrence formula for the exponential generating function of $f(n, n+k)$ for successive k . This proved suitable for machine computation. I cannot find anything similar for $s(n, n+k)$, but the method mentioned above takes its place.

We suppose henceforth that $k \geq 1$ and take any strong labelled $(n, n+k)$ digraph and suppose it to contain r points of degree (i.e. indegree + outdegree) greater than 2; we call these special points. The remaining (ordinary) points must each be of indegree one and outdegree one.

We have

$$2n+2k \geq 2(n-r)+3r$$

and so $1 \leq r \leq 2k$. Starting from any special point we follow a directed path through any number of ordinary points until

we again reach a special point. We call such a path a special path. If it begins and ends at the same special point, we call it an α - path. If the two special end-points are different and only one special path joins them in a given direction, we call the path a β - path. If more than one special path joins two different special points in the same direction, we call these paths γ - paths, unless one of them consists of a single edge, in which case we call it a δ - path. We suppose that the numbers of α -, β -, γ - and δ - paths in the digraph are a , b , c , d respectively.

If B is an ordinary point and AB , BC are directed edges, we say that we elide B if we remove B and replace AB , BC by the directed edge AC . An α - path must contain at least one ordinary point; we elide all its ordinary (interior) points but one. We elide all the ordinary points on every β - path and all but one on each γ - path. We now remove all the labels from the points. We are left with a strong unlabelled $(r+a+c, r+a+c+k)$ digraph, since each elision removes one point and one edge. Hence

$$2a+b+2c+d = r+a+c+k$$

and so

$$a+b+c+d = r+k \leq 3k. \quad (2.1)$$

Thus the number of possible resulting unlabelled non-isomorphic digraphs is bounded by a number depending on k . We can in theory construct all such digraphs (but not in practice for $k > 2$ without excessive labour).

We now take D any one of these unlabelled digraphs, suppose that the order of its automorphic group is g and determine how many different labelled $(n, n+k)$ digraphs reduce to D by the above process. If $n < r+a+c$, the number is zero. If

$$n = r+a+c+t, \quad t \geq 0,$$

we label the special paths and can distribute t new points among the α -, β - and γ - paths in just $y(t)$ ways, where $y(t)$ is the number of solutions of

$$m_1 + \dots + m_{a+b+c} = t$$

in non-negative integers, order being relevant. Then $y(t)$ is the coefficient of X^t in the expansion of $(1-X)^{-a-b-c}$ in ascending powers of X and so of X^n in

$$X^{r+a+c} (1-X)^{-a-b-c}.$$

We have thus constructed $y(n-r-a-c)$ digraphs with n points and $n+k$ edges. Let us label the n points in each digraph in all $n!$ possible ways, so that we now have a collection of $n!y(n-r-a-c)$ digraphs. But each digraph will occur just g times in the collection, so that exactly

$$j(n) = n!y(n-r-a-c)/g \quad (2.2)$$

different labelled $(n, n+k)$ digraphs reduce to D .

Next we observe that the coefficient of X^n in $X^u(1-X)^{-v}$ is

$$(n-u+1) \dots (n-u+v-1)/(v-1)!$$

This last expression vanishes for

$$n = u-1, u-2, \dots, u-v+1.$$

Hence $y(n-r-a-c)$ is a polynomial in n of degree $a+b+c-1$ which vanishes for

$$n = r+a+c-1, \dots, r-b+1.$$

Hence $j(n)$, defined by (2.2) with y a polynomial of degree $a+b+c-1$, is the number of non-isomorphic strong labelled $(n, n+k)$ digraphs reducible to D for all $n \geq r-b+1$.

Next we show that $r-b \leq k$. Elide all the remaining ordinary points in D so that we are left with a strongly connected multidigraph in which every special path is now

reduced to an edge. The α - paths are now loops; remove all a of them. Remove all but one of each set of parallel directed edges; one thus removes at least $\frac{1}{2}(c+d)$ edges. The resulting digraph has r points and is still strong. It must therefore have at least r edges and so not more than k edges have been removed. Hence $a + \frac{1}{2}(c+d) \leq k$ and so, by (2.1),

$$r - b = a + c + d - k \leq k.$$

Hence, for all $n > k$, we have just $j(n)$ strong labelled $(n, n+k)$ digraphs reducible to D . Summing over all the D , we have

$$s(n, n+k) = n! P_k(n) \quad (n > k),$$

where $P_k(n)$ is a polynomial whose degree is $\max(a+b+c-1)$.
 D

We know that $a+b+c \leq 3k$ and we can construct

a digraph D with $a+b+c = 3k$ as follows. Take $r = 2k$ and label the r special points $1, 2, \dots, r$. Let the $3k$ edges be $12, 23, \dots, r1$ and $32, 54, \dots, lr$. If we remove the labels, we have a digraph D with $a+b+c = b = 3k$, since each of the $3k$ edges is a β - path. This completes the proof of the first part of the theorem.

3. Calculation of $P_1(n)$. If $k = 1$, we have $r = 1$ or 2 and D can take any of the three forms in the figure.

The corresponding generating functions are

$$x^3(1-x)^{-2}, x^3(1-x)^{-2}, x^4(1-x)^{-3}$$

and the orders of the automorphic groups are 2, 1, 2

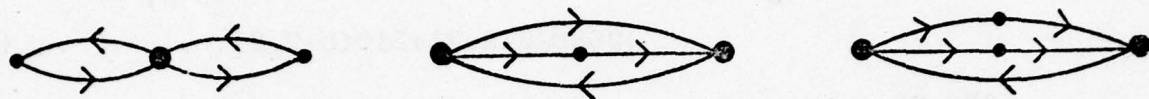
respectively. Hence the exponential generating function of $s(n, n+1)$ for $n > 1$ is

$$\frac{1}{2}x^3(1-x)^{-2} + x^3(1-x)^{-2} + \frac{1}{2}x^4(1-x)^{-3} = \frac{1}{2}x^3(3-2x)(1-x)^{-3}$$

and so we have

$$s(n, n+k) = \frac{1}{4}(n-2)(n+3)(n!),$$

$$2P_1(n) = \frac{1}{2}(n-2)(n+3) = B(n-2, 2) + 3B(n-2, 1).$$



4. Calculation of $P_k(n)$. If $k = 2$, we can determine the different forms of D with some labour; there are 30 of them. For each, we can determine $j(n)$ and g . Our result is that

$$24P_2(n) = 51B(n-2,5)+243B(n-2,4)+412B(n-2,3) \\ +268B(n-2,2)+24B(n-2,2). \quad (4.1)$$

and so

$$2880P_2(n) = (n-2)(51n^4+297n^3-271n^2-1937n-1020).$$

But this method becomes impossibly laborious for $k > 2$. It is simpler to compute $s(n, n+k)$ from the formulae of [2]. If we write

$$s_1 = 1, \quad s_n = s_n(Y) = \sum_{q=n}^{n(n-1)} s(n, q) Y^n$$

and define $\gamma_n = \gamma_n(Y)$ by $\gamma_1 = 1$ and

$$\gamma_n + \sum_{t=2}^{n-1} B(n, t) \gamma_t (1+Y)^{(n-1)(n-t)} = (1+Y)^{n(n-1)},$$

we have

$$s_n = \gamma_n + \sum_{t=1}^{n-1} B(n-1, t-1) s_t \gamma_{n-t}.$$

If we wish to determine $P_k(n)$ for all $k \leq K$, we need only compute the terms of γ_n and s_n up to Y^{5K} and can ignore all higher powers. We thus find the numerical values of

$$s(n, n+k) \quad (1 \leq k \leq K, \quad 2 \leq n \leq 4K)$$

and deduce those of

$$P_k(n) \quad (k+1 \leq n \leq 4K).$$

We write

$$P_k(n) = \sum_{m=1}^{3k-1} c_m B(n-k-1, m) + P_k(k+1) \quad (4.2)$$

and have

$$c_{h-1} = P_k(k+h) - P_k(k+1) - \sum_{m=1}^{h-2} c_m B(h-1, m),$$

giving c_1, \dots, c_{3k-1} in succession. All this is well adapted to machine computation. Dr. A. M. Murray was so kind as to obtain (4.1) in this way on the PDP 11/40 in the Computing Science Department of the University of Aberdeen. The form (4.2) for $P_k(n)$ is more convenient than a sum of powers, since $B(n, m)$ is always an integer.

References in Appendix 3

1. V. A. Liskovec, "On a recurrence method of counting graphs with labelled vertices", Soviet Math. Dokl. 10(1969), 242-256. (Dokl. Akad. S.S.S.R. 184(1969), No.6.)
2. E. M. Wright, "The number of strong digraphs", Bull. London Math. Soc. 3(1971), 348-350.
3. E. M. Wright, "The number of connected sparsely-edged graphs", submitted to the Journal of Graph Theory.

University of Aberdeen.

Appendix 4

Large cycles in large labelled graphs II

E.M. Wright†

University of Aberdeen.

(To appear in Math. Proc. Cambridge Phil. Soc.)

1. This paper is a sequel to [2]. The notation is the same and we assume the reader to be familiar with that paper. We there proved various theorems which established conditions under which "almost all" labelled (n, q) graphs contained a k -cycle, i.e. under which the proportion of all labelled (n, q) graphs which contained a k -cycle tended to 1 as $n \rightarrow \infty$. We here consider a different but related problem. If a graph contains a cycle of length k for every k such that $3 \leq k \leq k_0$, we say that the graph is pancyclic up to k_0 . We find conditions on k_0 , n and q such that almost all (n, q) graphs are pancyclic up to k_0 . If $k_0 = O(1)$, the result follows trivially from Theorem 3a of [1] (Theorem 1 of [2]), but if $k_0 \rightarrow \infty$ with n the matter is not so simple. We shall prove the following theorem.

† The research reported herein was supported by the European Research Office of the United States Army.

Theorem. If $q/n \rightarrow \infty$, $k_0^3 = o(n^2)$ and $k_0^3 = o(q^2/n)$ as $n \rightarrow \infty$, then almost all labelled (n, q) graphs are pancyclic up to k_0 .

The condition $q/n \rightarrow \infty$ is necessary, since otherwise only a finite proportion of the labelled (n, q) graphs contain a 3-cycle or, indeed, any given k -cycle with $k = O(1)$. I have no reason to suppose the other conditions necessary.

We no longer assume, as in [2], that $k \rightarrow \infty$ as $n \rightarrow \infty$. The proportion of labelled (n, q) graphs which lack one or more of the k -cycles for which $3 \leq k \leq k_0$ is not greater than

$$\sum_{k=3}^{k_0} (1-p) \leq \sum_{k=3}^{k_0} (\nu - 1) = S_1 + S_2,$$

where

$$S_1 = \sum_{k=3}^{k_0} (p - 1), \quad S_2 = \sum_{k=3}^{k_0} \{M(E)\}^{-1},$$

and we have only to show that $S_1 \rightarrow 0$ and $S_2 \rightarrow 0$ as $n \rightarrow \infty$.

By (3.2) of [2], we have

$$\frac{1}{M(E)} = \frac{2kN(N-1) \dots (N-k+1)}{n \dots (n-k+1)q \dots (q-k+1)} \leq k\epsilon^k,$$

where

$$\epsilon = n^2 / \{(n-k)(q-k)\} < Cn/q \rightarrow 0$$

as $n \rightarrow \infty$, since $k \leq k_0 = o(n)$

and $n = o(q)$. Hence

$$S_2 \leq \sum_{k \geq 3} k\epsilon^k \leq 3\epsilon^3(1-\epsilon)^{-2} \rightarrow 0$$

as $n \rightarrow \infty$.

We have to refine the argument at the bottom of p. 11 of [2] a little. We have

$$\begin{aligned} X &= \sum^n (\log \{1 - (s/q)\} - \log \{1 - (s/N)\}) \\ &\leq -\frac{k^2}{q} + \frac{2kr}{q} + O\left(\frac{k^3}{q^2}\right) + O\left(\frac{k^2}{n^2}\right) \end{aligned}$$

and so

$$\log \omega_r \leq r \log \alpha - (k^2/q) + O(k^3/q^2) + O(k^2/n^2).$$

As in [2],

$$\mu^{-1} K(n, k, r, R) \leq \frac{(2k^2)^R B(r-1, R-1)}{n^{2R-1} (n-k)^{r-R+1} R!}$$

and so

$$\rho \leq e^{-k^2/q} H \{1 + O(k^3/q^2) + O(k^2/n^2)\},$$

where

$$H = 1 + \sum_{r=1}^{k-2} \sum_{R=1}^r B(r-1, R-1) a^R b^{r-R} / R!$$

Now

$$b = n(n-1)e^{2k/q} / (2q(n-k)) = O(n/q) = o(1)$$

and so $H \leq \exp\{a/(1-b)\}$, as in [2]. Hence

$$\rho \leq \exp \left\{ \frac{a}{1-b} - \frac{k^2}{q} \right\} \left\{ 1 + O\left(\frac{k^3}{q^2}\right) + O\left(\frac{k^2}{n^2}\right) \right\}.$$

We have

$$a = \frac{k^2(n-1)}{qn}, \quad e^{2k/q} = \frac{k^2}{q} + O\left(\frac{k^2}{qn}\right) + O\left(\frac{k^3}{q^2}\right), \quad ab = O\left(\frac{nk^2}{q^2}\right),$$

and so

$$\frac{a}{1-b} - \frac{k^2}{q} = O\left(\frac{k^2}{qn}\right) + O\left(\frac{k^3}{q^2}\right) + O\left(\frac{nk^2}{q^2}\right).$$

But $k = o(n)$, $n = o(q)$, $q = O(n^2)$ and so

$$\rho \leq 1 + O(nk^2q^{-2}) + O(k^2n^{-2}).$$

Hence

$$\begin{aligned} \sum_{k=3}^{k_0} (\rho - 1) &= O(nq^{-2} + n^{-2}) \sum_{k=3}^{k_0} k^2 \\ &= O(k_0^3(nq^{-2} + n^{-2})) = o(1), \end{aligned}$$

by the conditions of the theorem.

References in Appendix 4

1. Erdős P. and Rényi, A. On the evolution of random graphs, Publ. Math. Inst. Hungarian Acad. Sci. 5(1960), 17-61.
2. Wright, E.M. Large cycles in large labelled graphs, Math. Proc. Cambridge Phil. Soc. 78(1975), 7.17.

Appendix 5

The evolution of unlabelled graphs

E. M. Wright †

(Submitted to the Journal of the London Math. Soc.)

1. Introduction. We are concerned here with (n,q) graphs. Such a graph has n points and q edges, but no loops or multiple edges. If the proportion of (n,q) graphs which do not have a property P tends to 0 as $n \rightarrow \infty$, we say that almost all (n,q) graphs have the property or that the typical (n,q) graph has the property P . We suppose n and q large and consider what are the properties of "almost all" unlabelled (n,q) graphs and how these evolve as q increases. Erdős and Renyi [2,3,4] studied the corresponding question for labelled (n,q) graphs. They showed that the evolution of the typical labelled (n,q) graph is fairly elaborate. I was able [8,9] to add a little to their results. In contrast, I shall show here that the evolution of the typical unlabelled (n,q) graph is relatively simple, that is, the behaviour can be described more simply and indeed is simpler, but the arguments required to establish this are more complicated than in the labelled case. Fortunately most of them have been deployed elsewhere [6].

† The research reported herein was supported by the European Research Office of the United States Army.

If a graph contains a cycle of length k for every k such that $3 \leq k \leq k_0$, we say that the graph is pancyclic up to k_0 . We write $N = n(n-1)/2$ and $\psi = \psi(q) = (2q/n) - \log n$. Again A is a positive number, not always the same at each occurrence and not dependent on n or q .

Theorem 1. If $q \rightarrow \infty$ as $n \rightarrow \infty$ but $\psi < A$, the typical unlabelled (n, q) graph for large q consists of one large connected component and otherwise only isolated points.
If $k_0^3 = o(q \log q)$, the connected component is pancyclic up to k_0 .

In [7] I proved that, if $\lim \psi \leq 0$ as $n \rightarrow \infty$, then almost all unlabelled (n, q) , graphs are disconnected. If $\psi > 0$, the proportion which are connected is asymptotic to $1 - e^{-\psi}$ and so tends to 1 if $\psi \rightarrow +\infty$. We shall prove the following theorem which gives some information about the pancyclic nature of the typical graph when $\psi \rightarrow \infty$.

Theorem 2. If $\psi \rightarrow +\infty$ as $n \rightarrow \infty$, $k_0^3 = o(q^2/n)$ and $k_0^3 = o(n^2)$, then the typical (n, q) graph is pancyclic up to k_0 .

2. Evolution of labelled graphs. It is interesting to compare the theorems I have just stated with the behaviour of the typical labelled (n,q) graph. Results about the latter are due to Erdős and Renyi [2,3,4] and [4] gives a particularly clear and detailed account (without proofs, which are given in [2] and [3]).

If $q = o(n)$, the typical labelled (n,q) graph consists of isolated trees and isolated points, trees of order k appearing first for small k as q increases. If $q \sim cn$, where $0 < c < \frac{1}{2}$, the largest component is a tree, but, for any k which is $O(1)$, there is a finite probability that the graph will contain a cycle of order k . If $q \sim cn$, where $c > \frac{1}{2}$, and $k_0^3 = o(n)$ the graph is pancyclic up to k_0 (see [8,9]). As q increases further, we have only fairly small trees outside a giant component. If $q \sim cn \log n$, where $c < \frac{1}{2}$, and $k_0^3 = o(n(\log n)^2)$, the typical graph consists of a large component pancyclic up to k_0 (by [9]) with isolated trees of order at most $[1/(2c)]$.

As q approaches $\frac{1}{2}n \log n$, the typical graph consists of one large component and $O(1)$ isolated nodes. If $-A < \psi < A$, the proportion of connected graphs is asymptotic to $\exp(-e^{-\psi})$. Finally, if $\psi \rightarrow +\infty$, the typical labelled graph is connected (by [2]) and is pancyclic up to k_0 (by [9]), provided that $k_0^3 = o(q^2 n^{-1})$ and $k_0^3 = o(n^2)$.

Thus, the evolution of the typical labelled graph is more complicated than that of the unlabelled graph. This arises from the differing effect of the permutation of the labels of the points on the numbers of different forms of graph for different sizes of q . We require the following lemma from the results for the labelled case.

Lemma 1. (Erdős and Renyi, [2], Lemma). If $-A < \psi < A$, then almost all labelled (n, q) graphs consist of one connected component and a number of isolated points.

3. The fundamental collection W. A collection of graphs may contain the same graph occurring more than once and the number of times it occurs is relevant. A set of graphs is a collection in which no graph occurs more than once. The addition of collections follows the rule one would expect, thus

$$\{a, a, b, c\} + \{a, b, c, c, c\} = \{a, a, a, b, b, c, c, c, c\} .$$

We consider the set T of unlabelled (n,q) graphs. We take every member of T, label its points from 1 to n and permute the labels in all n! ways. We denote the resulting collection by W and remark that

$$|W| = n! |T| \tag{3.1}$$

We write F_{π} for the set of labelled (n,q) graphs, each of which is invariant under the permutation π of the labels 1, ..., n of the points.

$$\text{Lemma 2: } W = \sum_{\pi} F(\pi), \tag{3.2}$$

where the sum is over all the permutations π of the n labels.

The so-called Burnside lemma, viz. that

$$n! |T| = \sum_{\pi} |F(\pi)|, \quad (3.3)$$

is an immediate consequence of Lemma 3 and (3.1). In fact, de Bruijn [1] gives a very short and clear account of the proof of Lemma 3 in the course of his proof of (3.3) and we refer the reader to [1] if the proof of Lemma 3 is not obvious to him.

4. Proof of Theorem 1. We say that a property P of an (n, q) graph is admissible if a labelled graph obtained by labelling an unlabelled graph has property P if and only if the unlabelled graph has property P . We see at once that almost all members of T have an admissible property P if and only if almost all members of W have that property.

For the present we suppose that

$$\frac{1}{2}n \leq 2q < n \log n + A n. \quad (4.1)$$

We now show that we can discard certain graphs from the collection W arranged in the sets listed on the right of (3.2) and still leave almost all of the members of W , i.e. if we write W'' for the collection of those graphs which we discard, and $W = W' + W''$, then

$$|W''| = o(|W|), |W'| \sim |W|.$$

To do this we refer to the calculations of [6].

We write $p = p(\pi)$ for the number of points whose labels are invariant under the permutation π and $F'(\pi)$ for the set of labelled (n, q) graphs which consist of a (p, q) graph on the p points invariant under π together with the remaining $n-p$ isolated points. If $P = \frac{1}{2}p(p-1)$, we have $|F'(\pi)| = P!/[q!(P-q)!]$. First, by Lemma 1 of [6] and the subsequent arguments, we see that we can discard all the graphs of those $F(\pi)$ for which

$$0 \leq p \leq M = [(q^4 n^2 \log n)^{\frac{1}{8}}].$$

Next, by Lemma 3 of [6], we can discard all the graphs of $F(\pi) - F'(\pi)$ for every π such that $M < p \leq n-2$.

As in [6], we write v for the positive number such that $v \log v = 2q$ and $V = [v]$. By the arguments of §§ 6-8 of [6], if $v \geq n$, we can discard the graphs of all $F'(\pi)$ except those for which

$$n\{1-(A/\log n)\} \leq p \leq n, \quad (4.2)$$

while, if $v < n$, we can discard the graphs of all $F'(\pi)$ except those for which

$$v\{1-(1/\log v)\} \leq p \leq \min(n, v\{1+(1/\log v)\}), \quad (4.3)$$

We are left then with W' , the collection of graphs formed by the summing of all those $F'(\pi)$ for which (4.2) or (4.3) is true, according as v is greater or less than n .

If

$$-A < (2q/p) - p \log p < A, \quad (4.4)$$

Lemma 1 applies and almost all the graphs in $F'(\pi)$ consist of one connected component and a number of isolated nodes. (This is clearly an admissible property). First, let us suppose that $v \geq n$, so that (4.2) is true, i.e.

$$p = n - (n\theta/\log n) \quad (0 \leq \theta \leq A)$$

and

$$2q = n \log n + \phi n \quad (0 \leq \phi \leq A).$$

It follows with a little calculation that

$$(2q/p) - \log p = \phi + \theta + o(1)$$

and so (4.4) is satisfied. Next, if $v < n$, we see from (4.3) that

$$p = v - (v \theta / \log v) \quad (-1 < \theta < 1)$$

and, since $2q = v \log v$,

$$(2q/p) - \log p = \theta + o(1),$$

so that (4.4) is satisfied. This completes the proof of the first part of Theorem 1.

To prove the second part we need the following lemma, proved in [9].

Lemma 3. If $q/p \rightarrow \infty$, $k_o^3 = o(p^2)$ and $k_o^3 = o(q^2/p)$ as $n \rightarrow \infty$, then almost all labelled (p, q) graphs are pancyclic up to k_o .

If $k_o^3 = o(q \log q)$ and (4.1) is true, with either (4.2) or (4.3) as appropriate, it is easy to deduce that the conditions of Lemma 3 are satisfied. Hence we have the second part of Theorem 1 under the condition (4.1).

We now remove the restriction $\frac{1}{2}n \leq q$ in (4.1).

If $q < \frac{1}{2}n$, every unlabelled (n, q) graph consists of a $(2q, q)$ graph together with $n-2q$ isolated points. The $(2q, q)$ graphs satisfy the condition (4.1) with $n = 2q$ and so the results of Theorem 1 are still true provided only that $q \rightarrow \infty$ with n .

5. Proof of Theorem 2. If $\psi(q)$ and $\psi(N-q)$ tend to infinity with n , the results of [5] tell us that almost all the members of W belong to $F(I)$, where I is the identity, i.e. consist of the set of labelled (n,q) graphs. Hence, for this range of q , Theorem 2 follows at once from Lemma 3.

There remains the case when $\psi(N-q) < A$, i.e. when $q > N - \frac{1}{2}n \log n - An$.

First let us consider the case when $N - \frac{1}{4}n \leq q \leq N$. Not more than $\frac{1}{4}n$ edges are "missing" from the complete (n,N) graph and so at least $\frac{1}{2}n$ of the points in the graph have degree $n-1$, i.e. each (n,q) graph contains a complete sub-graph on $[\frac{1}{2}n]$ points and so is pancyclic up to $[\frac{1}{2}n]$.

We need then only consider the case when

$$\frac{1}{4}n \leq N-q \leq \frac{1}{2}n \log n + An. \quad (5.1)$$

We now take $F'(\pi)$ to be the set of those labelled (n,q) graphs in which each of the $n-p$ points not invariant under π is of degree $n-1$. The sub-graph on the p invariant points has q' edges, where

$$\begin{aligned} q' &= q - \frac{1}{2}(n-p)(n-p-1) - p(n-p) \\ &= \frac{1}{2}p(p-1) - (N-q). \end{aligned}$$

We take v so that $v \log v = 2(N-q)$ and construct W' from the sets $F'(\pi)$ for which p satisfies (4.2) or (4.3) according as v is greater or less than n . Then almost all the members of W are in W' just as before, by the arguments and results of [6] .

If $p \geq \frac{1}{2}n$, then $q' > \frac{1}{4}p^2$ and so, by Lemma 3, provided $k_0^3 = o(p^2)$, i.e. $k_0^3 = o(n^2)$, almost all the graphs of $F'(\pi)$ are pancyclic up to k_0 , since this is true of the sub-graphs on the p invariant points. If $p < \frac{1}{2}n$, then $n-p \geq \frac{1}{2}n$ and so the complete sub-graph on the $n-p$ non-invariant points is pancyclic up to $n-p$, i.e. certainly up to $[\frac{1}{2}n]$. Hence, in this case, every graph of $F'(\pi)$ is pancyclic up to $[\frac{1}{2}n]$.

It follows that, when (5.1) is true, almost all the graphs of W' , and so of W , are pancyclic up to k_0 , where $k_0^3 = o(n^2)$. This completes the proof of Theorem 2.

References in Appendix 5

1. N. G. de Bruijn, Chap. 5. Applied combinatorial mathematics, E. F. Beckenback (ed.), New York 1964.
2. P. Erdős and A. Rényi, On random graphs I, Publ. Math. Debrecen 6(1959), 290-297.
3. P. Erdős and A. Rényi, On the evolution of random graphs, Magyar Tud. Akad. Mat. Kut. Int. Közl. 5(1960), 17-61.
4. P. Erdős and A. Rényi, On the evolution of random graphs, Bull. Inst. Internat. Statistic. 38(4) (1961), Tokyo, 343-347.
5. E. M. Wright, Graphs on unlabelled nodes with a given number of edges, Acta Math. 126(1971), 1-9.
6. E. M. Wright, Graphs on unlabelled nodes with a large number of edges, Proc. London Math. Soc. 28(1974), 577-594.
7. E. M. Wright, The probability of connectedness of a large unlabelled graph, Journ. London Math. Soc. (2), 11(1975), 13-16.
8. E. M. Wright, Large cycles in large labelled graphs, Math. Proc. Cambridge Phil. Soc. 78(1975), 7-17.
9. E. M. Wright, Large cycles in large labelled graphs II, to appear in Math. Proc. Cambridge Phil. Soc.

University of Aberdeen.

Appendix 6

The degrees of nodes in almost all graphs.

E. M. Wright

(Submitted to Math. Annalen.)

G is a graph on n labelled nodes x_1, \dots, x_n . Every pair of different nodes is or is not joined by just one edge. There are no loops. We write $N = \frac{1}{2}n(n-1)$ and consider the set of 2^N different graphs G . If the proportion of these graphs which do not have a property Q tends to 0 as $n \rightarrow \infty$, we say that almost every graph G has the property Q . The positive number C , not always the same at each occurrence, is independent of n ; the numbers C_1, C_2, C_3, C_4 each denote the same fixed number throughout. We write $d(x_i)$ for the degree of the node x_i and $B(h,k)$ for the binomial coefficient $h!/[k!(h-k)!]$. We shall prove two theorems.

Theorem 1. For some C_1 and all i ($1 \leq i \leq n$)
we have

$$\frac{1}{2}n - C_1(n \log n)^{\frac{1}{2}} < d(x_i) < \frac{1}{2}n + C_1(n \log n)^{\frac{1}{2}}$$

in almost every graph G .

† The research reported herein was supported by the European Research Office of the United States Army.

Theorem 2. For some C_2 and C_3 , there are, in almost every graph G , more than $C_3 n/\log n$ nodes x_i such that $d(x_i) < \frac{1}{2}n - C_2 n^{\frac{1}{2}}$ and more than $C_3 n/\log n$ nodes x_i such that $d(x_i) > \frac{1}{2}n + C_2 n^{\frac{1}{2}}$.

The two theorems show that, in almost all G , the degrees of the nodes are all fairly close to $\frac{1}{2}n$ but a substantial number are not too close. It would be interesting to find improvements of these results or to show that, in some sense or other, any of them are best possible. It is enough to prove the first part of each theorem. The second part follows in each case by considering the complementary graphs.

We write $m = \lfloor \frac{1}{2}(n-1) \rfloor$ and u for any integer such that $0 < u < m$. The probability that $d(x_1) = j$ is $2^{1-n} B(n-1, j)$ and so the probability that $d(x_1) \leq m - u$ is

$$P_u = 2^{1-n} \sum_{j=0}^{m-u} B(n-1, j) \leq m 2^{1-n} B(n-1, m-u).$$

If we now put $v = \lfloor C_1 (n \log n)^{\frac{1}{2}} \rfloor$, where $C_1^2 > \frac{3}{4}$, a routine calculation shows us that

$$B(n-1, m-v)/B(n-1, m) < C n^{-\frac{3}{2}-C}$$

and, by Stirling's Theorem,

$$B(n-1, m) < Cn^{-\frac{1}{2}} 2^n, \quad (1)$$

so that

$$P_v < Cn 2^{1-n} B(n-1, m-v) \leq Cn^{-1-C}.$$

Hence the probability that any of the nodes

x_1, \dots, x_n has degree less than $m - C_1(n \log n)^{\frac{1}{2}}$ is less than Cn^{-C} , which tends to 0 as $n \rightarrow \infty$.

Theorem 1 follows.

Next

$$\sum_{j=0}^{m-1} B(n-1, j) + aB(n-1, m) = 2^{n-2},$$

where $a = \frac{1}{2}$ or 1 according as n is odd or even.

Again, for any integer h , we have

$B(n-1, m-h) \leq B(n-1, m)$. Hence,

$$P_u \geq \frac{1}{2} - 2^{1-n} u B(n-1, m) \geq \frac{1}{2} - Cn^{-\frac{1}{2}} \quad (2)$$

by (1).

We write α_j for the probability that

$d(x_j) < \frac{1}{2}n - u$, given that $d(x_i) \geq \frac{1}{2}n - u$ for all i such that $1 \leq i < j$. Then the probability that at least one of $d(x_i)$ is less than $\frac{1}{2}n - u$ (where $1 \leq i \leq k$) is

$$\bar{\Phi} = \bar{\Phi}(\alpha_1, \dots, \alpha_k) = 1 - \prod_{j=1}^k (1 - \alpha_j)$$

We write β_j for the probability that $d(x_j) < \frac{1}{2}n - u - j + 1$ in G_j , the graph obtained from G by deleting the nodes x_1, \dots, x_{j-1} and all edges adjacent to them. From (2) with $n-j+1$ for n , we see that, provided $j \leq k$,

$$\beta_j \geq \frac{1}{2} - C(n-j)^{-\frac{1}{2}}(2u+j) \geq 1 - \delta$$

where

$$\delta = \frac{1}{2}\{1 + C(n-k)^{-\frac{1}{2}}(2u+k)\} < C_4 < 1$$

if $k < n^{\frac{1}{4}}$ and $u < C_2 n^{\frac{1}{2}}$ for a suitable C_2 . But it is clear that $\alpha_j \geq \beta_j$ and so $\alpha_j \geq 1 - \delta$.

The polynomial Φ increases with each α_i and so

$$\begin{aligned} \Phi(\alpha_1, \dots, \alpha_k) &\geq \Phi(1-\delta, \dots, 1-\delta) \\ &= 1 - \delta^k \geq 1 - C_4^k, \end{aligned}$$

that is, the probability that the degree of one of x_1, \dots, x_k is less than $m-u$ is greater than $1 - C_4^k$.

Hence, if we select, from all possible G , those for which $d(x_i) < m - u$ for one at least of x_1, \dots, x_k , we have at least $(1 - C_4^k) 2^{n(n-1)/2}$ graphs G . The restriction $d(x_i) < m - u$ for any $i \leq k$ cannot increase the probability that a particular edge is present and hence cannot decrease the

probability that $d(x) < m - u$ for any $x_i (i > k)$.

Hence we have at least

$$(1 - C_4^k)^2 2^{n(n-1)/2}$$

graphs G for which $d(x_i) < m - u$ for at least one i for which $1 \leq i \leq k$ and for at least one i for which $k + 1 \leq i \leq 2k$. Continuing thus, we find that, provided $sk \leq n$, the probability that there are s different nodes each with degree less than $m - u$ is at least

$$(1 - C_4^k)^s \geq 1 - sC_4^k \rightarrow 1$$

as $n \rightarrow \infty$, provided we choose $k = \lceil (\log n) / \log(1/C_4) \rceil$ and $s = \lfloor n/k \rfloor$. Theorem 2 follows.

We remark that both theorems are true for unlabelled graphs. Almost all labelled graphs satisfy Theorem 1 and so have their number of edges between $\frac{1}{4}n^2 \pm Cn^{\frac{3}{2}}(\log n)^{\frac{1}{2}}$. Hence, by [2], they are asymmetric, i.e. fall into equivalence classes under the symmetric group, each of which classes corresponds to a different unlabelled graph. The number of such classes is asymptotic to $2^N/n!$, i.e. by [1], asymptotic to the total number of unlabelled graphs on n nodes. Hence the results of Theorem 1 and Theorem 2 remain true for almost every unlabelled graph.

It is an open question whether either theorem is in some sense best possible or, if not, whether it can be improved.

References

1. E. M. Wright, The number of graphs on many unlabelled nodes, Math. Ann. 183(1969), 250-253.
2. E. M. Wright, Asymmetric and symmetric graphs, Glasgow Math. Journ. 15(1974), 69-73.

University of Aberdeen.

References to Literature in Report

1. A. Cayley, A theorem on trees, Quart. Journ. Math. 23 (1889), 376-378.
2. P. Erdős and A. Rényi, On random graphs I, Publ. Math. Debrecen 6(1959), 290-297.
3. P. Erdős and A. Rényi, On the evolution of random graphs, Magyar Tud. Akad. Mat. Kut. Int. Közl. 5(1960), 17-61.
4. P. Erdős and A. Rényi, On the evolution of random graphs, Bull. Inst. Internat. Statistic. 38(4) (1961), 343-347.
5. E. N. Gilbert, Enumeration of labelled graphs, Canadian Journ. Math. 8(1956), 405-411.
6. V. A. Liskovec, On a recurrence method of counting graphs with labelled vertices, Soviet Math. Dokl. 10(1969), 242-256. (Dokl. Akad. Nauk SSSR 184(1969), No.6).
7. J. W. Moon, Various proofs of Cayley's formula for counting trees, A seminar on graph theory (ed. F. Harary, New York 1967), 70-78.
8. A. Rényi, On connected graphs I, Publ. Math. Inst. Hungarian Acad. Sci. 4(1959), 385-388.
9. R. J. Riddell and G. E. Uhlenbeck, On the theory of the virial development of the equation of state of monoatomic gases, Journ. Chem. Phys. 21(1953), 2056-2064.
10. H. N. V. Temperley, On the enumeration of Mayer cluster integrals, Proc. Phys. Soc. 72(1959), 1141-1144.
11. E. M. Wright, Asymptotic enumeration of connected graphs, Proc. Roy. Soc. Edinburgh A 68(1970), 298-308.
12. E. M. Wright, The number of strong digraphs, Bull. London Math. Soc. 3(1971), 348-350.
13. E. M. Wright, Graphs on unlabelled nodes with a large number of edges, Proc. London Math. Soc. 28(1974), 577-594.
14. E. M. Wright, Large cycles in large labelled graphs, Math. Proc. Cambridge Phil. Soc. 78(1975), 7-17.